**Stat and Data Analysis Activity 4.2**

1. How well do you know the ages of some of the stars and celebrities that you hear about and see on a regular basis? You are to complete the following table with what you believe to be the person’s age. FILL IN THE “YOUR GUESS” COLUMN. After you are done, you will be given their real ages.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NAME | Your guess | Actual Age | Residual | Residual Squared |
| Barrack Obama |  |  |  |  |
| Bruce Willis |  |  |  |  |
| Lindsay Lohan |  |  |  |  |
| Kristen Stewart |  |  |  |  |
| Sandra Bullock |  |  |  |  |
| Lady Gaga |  |  |  |  |
| Hugh Hefner |  |  |  |  |
| Al Pacino |  |  |  |  |
| Brad Pitt |  |  |  |  |
| Angelina Jolie |  |  |  |  |
| Mr. Davidheiser |  |  |  |  |

2. To see how well you have done, we will analyze the differences between the person’s actual age and your guess. On a piece of graph paper create a scatterplot with YOUR GUESS as the explanatory variable and the ACTUAL AGE as the response variable.

3. Now sketch the line Y = X. What does it mean if your points fall on the line?

4. If your point is NOT on the line Y = X, then draw a vertical line from your point to the line. The distance of each of these segments is a ***residual*** for that guess. To find the direction of the residual, subtract ACTUAL AGE by YOUR GUESS. (ACTUAL AGE– YOUR GUESS). Some will be positive and some negative. List these values in the **RESIDUAL** column of the table above.

5. Did you tend to *underestimate* (points mostly above the line)*, overestimate* (points are mostly below the line) or *bounce around* the line?

6. Find the sum of the residual column: \_\_\_\_\_\_. Does your answer tell you how well you guessed the ages overall? \_\_\_\_\_\_ Explain why or why not.

7. **Square** each residual value to force it to become positive. Write these in the RESIDUAL SQUARED column.

8. This can be represented on the graph by creating squares from the point to the line using the residual as a side length. (Check out p. 171 in the book for an example.). Do this on your sctterplot.

9. Find the sum of the squared residual values: \_\_\_\_\_\_\_\_\_\_\_\_

**Read over pp. 169-172**

**Answer the following**

1. What is a regression line?
2. What is a regression line used for?
3. What was the regression line used for in example 4.6?
4. What was the regression line used for in example 4.7?
5. What were the correlations for Example 4.6 and Example 4.7? What do they tell us about the regression lines for each example?
6. When drawing a regression line we want the line to be close the points in what direction? Why?
7. What is the most common method of regression?
8. Define Least-Squares Regression Line (LSRL).
9. What does the LSRL try to minimize?
10. What is the form of the equation for the LSRL?
11. What two numbers are necessary for the equation of the LSRL?
12. In Example 4.8 what is the equation for the LSRL?
13. What was the slope? What was the interpretation of slope in the context of the problem?
14. What was the y-intercept? What is the interpretation of the y-intercept? Does it have an important meaning?
15. If you want to add the LSRL to a scatterplot what must you do?



